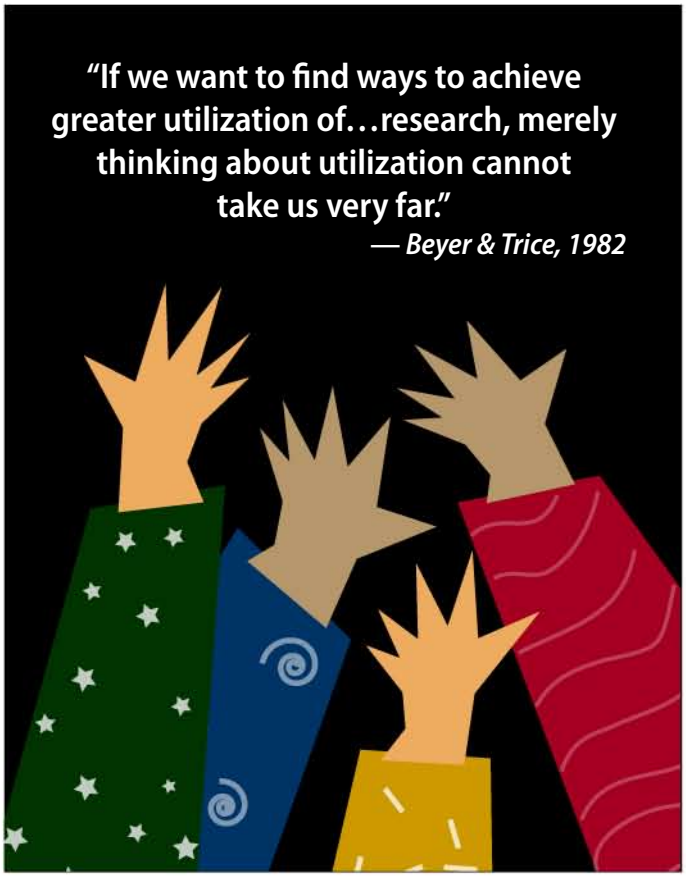
Chapter 1

Introduction

- Review Methods
- An Implementation Headset
- Implementation Defined
- Paper Implementation



Introduction

It has been well documented in many disciplines that major gaps exist between what is known as effective practices (i.e., theory and science) and what is actually done (i.e., policy and practice).

Background & Purpose

"The ideas embodied in innovative social programs are not self-executing."

—Petersilia, 1990

In the past few years several major reports highlighted the gap between our knowledge of effective treatments and services currently being received by consumers. These reports agree that we know much about interventions that are effective but make little use of them to help achieve important behavioral health outcomes for children, families, and adults nationally. This theme is repeated in reports by the Surgeon General (United States Department of Health and Human Services, 1999; 2001), the National Institute of Mental Health [NIMH] National Advisory Mental Health Council Workgroup on Child and Adolescent Mental Health Intervention Development and Deployment (2001), Bernfeld, Farrington, & Leschied (2001), Institute of Medicine (2001), and the President's New Freedom Commission on Mental Health (2003). The authors call for applied research to better understand service delivery processes and contextual factors to improve the efficiency and effectiveness of program implementation at local, state, and national levels.

Our understanding of how to develop and evaluate evidence-based intervention programs has been furthered by on-going efforts to research and refine programs and practices, to define "evidence bases" (e.g., Burns, 2000; Chambless & Ollendick, 2001; Lonigan, Elbert, & Johnson, 1998; Odom, et al., 2003), and to designate and catalogue "evidence-based programs or practices" (e.g., the National Registry of Evidence-based Practices and Programs, Substance Abuse and Mental Health Services Administration, n.d.; Colorado Blueprints for Violence Prevention, Mihalic, Fagan, Irwin, Ballard, & Elliott, 2004). However, the factors involved in successful implementation of these programs are not as well understood (Backer, 1992; Chase, 1979; Leonard-Barton & Kraus, 1985; Reppucci & Saunders, 1974; Rogers, 1983, 1995; Shadish, 1984; Stolz, 1981; Weisz, Donenberg, Han, & Weiss, 1995). Current views of implementation are based on the scholarly foundations prepared by Pressman & Wildavsky's (1973) study of policy implementation, Havelock & Havelock's (1973) classic curriculum for training change agents, and Rogers' (1983; 1995) series of analyses of factors influencing decisions to choose a given innovation. These foundations were tested and further informed by the experience base generated by pioneering attempts to implement Fairweather Lodges (Fairweather, Sanders, & Tornatzky, 1974) and National Follow-Through education models (Stivers & Ramp, 1984; Walker, Hops, & Greenwood, 1984), among others. Petersilia (1990) concluded that, "The ideas embodied in innovative social programs are not self-executing." Instead, what is needed is an "implementation perspective on innovation—an approach that views postadoption events as crucial and focuses on the actions of those who convert it into practice as the key to success or failure" (p. 129). Based on their years of experience, Taylor, Nelson, & Adelman (1999) stated, "Those who set out to change schools and schooling are confronted with two enormous tasks. The first is to develop prototypes. The second involves large scale replication. One without the other is insufficient. Yet considerably more attention is paid to developing and validating prototypes than to delineating and testing scale-up processes. Clearly, it is time to correct this deficiency." (p. 322). Gendreau, Goggin, & Smith (1999) added that, "we cannot afford to continue dealing with the business of program implementation and related technology transfer topics in a cavalier fashion" (p. 185).

The purpose of this monograph is to describe the results of a far-reaching review of the implementation literature. There is broad agreement that implementation is a decidedly complex endeavor, more complex than the policies, programs, procedures, techniques, or technologies that are the subject of the implementation efforts. Every aspect

of implementation is fraught with difficulty, from system transformation to changing service provider behavior and restructuring organizational contexts. Given the importance of implementation, the purpose of this review is to create a topographical map of implementation as seen through evaluations of factors related to implementation attempts. It is not an attempt to be exhaustive. Some literature reviews have very exacting criteria and review procedures, a style well-suited to areas of well-developed knowledge. With respect to implementation, there is no agreed-upon set of terms, there are few organized approaches to executing and evaluating implementation practices and outcomes, and good research designs are difficult when there are "too many variables and too few cases" (Goggin, 1986). Given the state of the field, the goal was to "review loosely" to capture meaning, detect relationships among components, and help further the development of the practice and science of implementation.

The remainder of this introduction sets the stage for reading the monograph. There is an overview of the review methods in order to provide the reader with a context for evaluating the face validity of the review in terms of scope, findings, and frameworks. This is followed by an orientation to implementation as distinct from program development and a definition of implementation.

Review Methods

The goal of this literature review is to synthesize research in the area of implementation as well as to determine what is known about relevant components and conditions of implementation. Search strategies were developed by the research team as an iterative process in consultation with the Louis de la Parte Florida Mental Health Institute (FMHI) University of South Florida librarian. The research team began the literature searching process by establishing guidelines for citation retrieval. The following citation retrieval criteria were used to select reports, books, and published and unpublished article citations for preliminary review:

- published in English no earlier than 1970,
- the title or abstract contained one or more of the search terms, and
- an empirical study, meta-analysis, or literature review.

Literature with any data (quantitative or qualitative) and any design (surveys to high quality randomized group designs or within subject designs) in any domain (including agriculture, business, child welfare, engineering, health, juvenile justice, manufacturing, medicine, mental health, nursing, and social services) was eligible for inclusion.

Databases searched included PsycINFO, Medline, Sociological Abstracts, CINAHL, Emerald, JSTOR, Project Muse, Current Contents, and Web of Science. Once the research team had completed the literature search, nearly 2,000 citations were retrieved and entered into an EndNote database. The principal investigators then proceeded to pare down the list by reading the titles and abstracts using the same guidelines for citation retrieval (full details are provided in Appendix A). The remaining citations (N = 1,054) were retrieved for full-text review and content analysis. The review team developed a data extraction tool called the *article summary* to record pertinent information from each document reviewed. The article summary covered several aspects including: the research domain, topic or purpose of the article, methods, results and findings, codes or stages of implementation as defined by the codebook, selected quotations, selected references, and memos or notes made by the reviewer about the article.

Full text reviews were completed by one of the five review team members. Each team member was asked to make note of any particularly noteworthy or "significant" implementation articles in the memo section of the article summary if it met one of the following three criteria: (1) well-designed experimental evaluations of implementation factors, (2) careful reviews the implementation literature, or (3) well-thought-out but more theoretical discussions of implementation factors. For example, "significant" articles included literature describing group or within-subject experimental designs, meta-analyses, or literature reviews pertaining to specific implementation factors; literature describing useful frameworks or theoretical summaries; or qualitative analyses of specific implementation efforts. Literature that focused on author-generated surveys of those involved in implementation efforts, focused on interventions and only provided incomplete de-

Given the state of the field, the goal was to "review loosely" to capture meaning, detect relationships among components, and help further the development of the practice and science of implementation.

scriptions of implementation factors, or primarily presented the opinions of the authors were not included as "significant" articles.

After reading the full text, about 30% of the 1,054 articles were dropped from the review. Most often, deletions occurred when implementation was mentioned in the title or abstract but was not evaluated in any way in the article itself (i.e., was not "an empirical study, meta-analysis, or review"). Once the full text review was completed, 743 articles remained, about half (377) of which were identified as significant implementation articles. Of these, 22 articles reported the results of experimental analyses (randomized group or within subject designs) or meta-analyses of implementation variables. Article summaries were sorted into content areas by searching across articles for the codes described in the codebook (see Appendix B). The principal investigators then proceeded to review each area for common implementation themes and patterns.

The review was challenging due to the lack of well-defined terms. Diffusion, dissemination, and implementation sometimes referred to the same general constructs and, at other times, quite different meanings were ascribed to the same terms. For example, "implementation" sometimes means "used" in a general sense or "put into effect" with specific reference to a program or practice. At other times it referred to a set of methods to purposefully help others make use of a program or practice on a broad scale. Similarly, coaching, supervision, academic detailing, and on-the-job teaching were used to describe similar activities. Are the "implementers" the ones teaching or the ones being taught? The answer is, it depends on the author. We have created our own lexicon with definitions (see Appendix A and B) in the text to help guide the reader through this monograph and to reduce confusion. The lack of common definitions and the lack of journals specifically oriented to implementation research probably reflect the poorly developed state of the field.

The lack of common definitions and the lack of journals specifically oriented to implementation research probably reflect the poorly developed state of the field.

An Implementation Headset

It is important to have an "implementation headset" while reading this monograph. From an implementation point of view, there are always two important aspects of every research study, demonstration project, or attempted intervention. In each study, there are intervention processes and outcomes and there are implementation processes and outcomes. When implementing evidence-based practices and programs, Blase, Fixsen, & Phillips (1984) discussed the need to discriminate implementation outcomes (Are they doing the program as intended?) from effectiveness outcomes (Yes, they are, and it is/is not resulting in good outcomes.). Only when effective practices and programs are fully implemented should we expect positive outcomes (Bernfeld, 2001; Fixsen & Blase, 1993; Institute of Medicine, 2001; Washington State Institute for Public Policy, 2002).

So far, as the wave of interest in evidence-based practices and programs has swept across human services, the nature of the evidence about interventions has received the preponderance of attention from researchers and policy makers. As Kitson, Harvey, & McCormack (1998) stated, "... the investment in developing structures to ensure gold standard research evidence has yet to be matched by equal investment in ways of elucidating how organizations change cultures or use different techniques to manage the change process" (p 157). From an implementation point of view, doing more and better research on a program or practice itself does not lead to more successful implementation. A series of meta-analyses and detailed assessments of the strength of research findings for certain practices and programs may help a consumer, agency, or community select a program. However, more data on program outcomes will not help implement that program. Implementation is an entirely different enterprise. Thus, an intervention must be well defined and carefully evaluated with regard to its effects on its intended consumers (children, families, adults). Likewise, implementation of an intervention must be well defined and carefully evaluated with regard to its effects on its intended consumers (practitioners, managers, organizations, systems).

An implementation headset also is critical for understanding and interpreting data from outcome

studies. Rossi & Freeman (1985) identified three ways in which inadequate measures of program implementation may lead to an incorrect conclusion that an intervention is ineffective. First, no treatment or too little treatment is provided; second the wrong treatment is provided; and third, the treatment is nonstandard, uncontrolled, or varies across the target population. Dobson & Cook (1980) described "type III" (type three) errors. That is, evaluating a program that was described but not implemented. In their analysis of a program for ex-offenders, they found only 1 in 20 consumers actually received the program as described in the methods section. Thus, the outcome data could not be attributed to the program as described. Feldman, Caplinger, & Wodarski (1983) found that apparent findings of no differences among groups were explained by measuring the application of the independent variables. Those youths who were in groups whose leaders skillfully followed the protocol had better outcomes.

Outcome interpretation is further compromised when control groups utilize the components of the evidence-based program or practice, or, if the experimental programs fail to implement key aspects of the intervention. In studies of one evidence-based program (Assertive Community Treatment or ACT; Bond, Evans, Salyers, Williams, & Kim, 2000) it was found in one case that a control site had incorporated many ACT principles (McHugo, Drake, Teague, & Xie, 1999), while in another that the experimental sites had implemented fewer aspects of the ACT model than expected (Bond, Miller, Krumweid, & Ward, 1988). Dane & Schneider (1998) conducted a literature review of prevention programs published between 1980 and 1994. They found that only 39 (24%) of 162 outcome studies documented the implementation of the independent variables (i.e., fidelity) and only 13 used a measure of fidelity as a variable when analyzing the results. They also noted that the amount of documentation of fidelity found in their review (24%), "compared to the 20% found by Peterson, et al. (1982) in 539 experimental studies published from 1968 to 1980 in the *Journal of Applied Behavior Analysis,* the 18.1% found by Moncher and Prinz (1991) in 359 treatment outcome studies published in clinical psychology, psychiatry, behavior therapy, and family therapy journals from 1980 to 1988, the 6% found by Rogers-Weise in 88 group-design parent training studies published from 1975 to 1990, and the 14.9% noted by Gresham et al. (1993) in evaluations of behaviorally based interventions published from 1980 to 1990" (p. 41). Dane & Schneider (1998) concluded that, "A reorganization of research priorities is needed to facilitate less confounded, better quality evaluations of preventive interventions" (p. 42).

Thus, implementation variables are not synonymous with those involved in interventions and implementation outcomes are important to measure, analyze, and report when attempting to interpret research findings or broad scale applications (Bernfeld, 2001; Blase et al., 1984; Dusenbury, Brannigan, Falco, & Hansen, 2003; Forsetlund, Talseth, Bradley, Nordheim, & Bjorndal, 2003; Goodman, 2000; Mowbray, Holter, Teague, & Bybee, 2003; Rychetnik, Frommer, Hawe, & Shiell, 2002).

Implementation Defined

What is "implementation?" For the purposes of this review, implementation is defined as a specified set of activities designed to put into practice an activity or program of known dimensions. According to this definition, implementation processes are purposeful and are described in sufficient detail such that independent observers can detect the presence and strength of the "specific set of activities" related to implementation. In addition, the activity or program being implemented is described in sufficient detail so that independent observers can detect its presence and strength. As noted earlier, when thinking about implementation the observer must be aware of two sets of activities (intervention-level activity and implementation-level activity) and two sets of outcomes (intervention outcomes and implementation outcomes).

The view becomes a bit more complicated when implementation-savvy researchers talk about implementation-related "interventions" with community leaders, agency directors, supervisors, practitioners, policy makers, and funders. For purposes of this monograph, we will use "interventions" to mean treatment or prevention efforts at the consumer level and "implementation" to mean efforts to incorporate a program or practice at the

For the purposes of this review:
Implementation is defined as a specified set of activities designed to put into practice an activity or program of known dimensions.

community, agency, or practitioner levels. Also, it is common to read about "implementation" of a program or practice as if it were an accomplished fact when the context of the statement makes it clear that some process (more or less clearly described) had been put in place to attempt the implementation of that program or practice (e.g., funding, policy mandate). When faced with the realities of human services, implementation outcomes should not be assumed any more than intervention outcomes are assumed.

Degrees of Implementation

During the course of the review, it was noted that various authors discussed the purposes and outcomes of implementation attempts in different ways (Goggin, 1986). The purposes and outcomes of implementation might be categorized as:

Paper implementation means putting into place new policies and procedures (the "recorded theory of change," Hernandez & Hodges, 2003) with the adoption of an innovation as the rationale for the policies and procedures. One estimate was that 80-90% of the people-dependent innovations in business stop at paper implementation (Rogers, 2002). Westphal, Gulati, & Shortell (1997) found in their survey of businesses that, "If organizations can minimize evaluation and inspection of their internal operations by external constituents through adoption alone, they may neglect implementation altogether, decoupling operational routines from formally adopted programs." (p. 371). Thus, paper implementation may be especially prevalent when outside groups are monitoring compliance (e.g., for accreditation) and much of the monitoring focuses on the paper trail. It is clear that paperwork in file cabinets plus manuals on shelves do not equal putting innovations into practice with benefits to consumers.

Process implementation means putting new operating procedures in place to conduct training workshops, provide supervision, change information reporting forms, and so on (the "expressed theory of change" and "active theory of change," Hernandez & Hodges, 2003) with the adoption of an innovation as the rationale for the procedures. The activities related to an innovation are occurring, events are being counted, and innovation-related languages are adopted.

However, not much of what goes on is necessarily functionally related to the new practice. Training might consist of merely didactic orientation to the new practice or program, supervision might be unrelated to and uninformed by what was taught in training, information might be collected and stored without affecting decision making, and the terms used in the new language may be devoid of operational meaning and impact. In business, this form of implementation has been called the Fallacy of Programmatic Change. That is, the belief that promulgating organizational mission statements, "corporate culture" programs, training courses, or quality circles will transform organizations and that employee behavior is changed simply by altering a company's formal structure and systems (Beer, Eisenstat, & Spector, 1990). It is clear that the trappings of evidence-based practices and programs plus lip service do not equal putting innovations into practice with benefits to consumers.

Performance implementation means putting procedures and processes in place in such a way that the identified functional components of change are used with good effect for consumers (the "integrated theory of change," Hernandez & Hodges, 2003; Paine, Bellamy, & Wilcox, 1984). It appears that implementation that produces actual benefits to consumers, organizations, and systems requires more careful and thoughtful efforts as described by the authors reviewed in this monograph.